UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

MASTER FILE

March 21, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES R-29

MEMORANDUM FOR Dennis Stoudt
Assistant Division Chief, Processing Systems
Decennial Systems and Contracts Management Office

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Peter P. Davis PPD
Sample Design Team
Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation Survey: Reduction Specification

I INTRODUCTION

This memorandum describes the selection of the Accuracy and Coverage Evaluation (A.C.E.) reduction sample. The A.C.E. reduction is the sampling operation that links the original Integrated Coverage Measurement (ICM) Survey sampling plan to the A.C.E. sampling plan. The A.C.E. reduction is the first of several operations that will reduce the number of housing units (HUs) from the nearly two million HUs in independent listing to the approximately 300,000 HUs that will be sent for interview. Block clusters were selected for A.C.E. independent listing under the previously planned 750,000 HU ICM design. (See References 1, 2, 3, 4, and 5.) Since not all of the listed block clusters are required for A.C.E., the reduction will subsample those clusters, with the selected clusters continuing on in A.C.E. operations. The A.C.E. reduction is a subsample of medium and large block clusters in the 50 states and the District of Columbia. Small block clusters and block clusters on American Indian Reservations (AIR) and in Puerto Rico are not eligible for this reduction. The two remaining sampling operations are small block cluster subsampling and large block cluster subsampling.

In small block cluster subsampling, the 5,000 small block clusters selected in the A.C.E. listing sample will be subsampled following the keying of the independent listing books. As a general rule, if a cluster is larger than expected, then the cluster will be retained at a higher rate. This stage of sampling should not have a significant impact on the interview

sample size but is critical for controlling the size of the cluster weights and for maximizing field interviewing efficiency.

The final stage of selection is large block cluster subsampling. When a cluster has more than 80 listed HUs, segments of adjacent HUs are formed and a subsample of segments is selected from the cluster. This produces the final A.C.E. interview sample.

The sections of this specification are ordered as follows:

- Section II: Overview
- Section III: Definitions
- Section IV: Assumptions
- Section V: Input Files
- Section VI: Output Files
- Section VII: Reduction Process
- Section VIII: References

Note that the A.C.E. reduction operation is complete. This specification reflects the reduction as it was actually implemented. The results of the reduction are documented in reference 6, and a contingency plan that addresses the delay of an updated census address list is documented in reference 7. This final specification includes changes to the original working draft that were required due to issues that arose in testing or production. The final specification retains the flexible approach to the reduction that was presented in the working draft. This approach was required since system development had to begin before a final sample design could be determined. Therefore, the reduction parameters are treated as variables in this specification even though in the final sample design they may have ended up as constants. For example, all medium stratum jumper clusters, those that moved from the medium sampling stratum to the large, were retained in the reduction sample even though their probability of selection is treated as a variable in this specification. In general, this final specification is nearly identical to the original working draft.

Any comments or questions should be directed to Pete Davis (301-457-8322), Jim Farber (301-457-4282), or Debbie Fenstermaker (301-457-4195) of the Decennial Statistical Studies Division (DSSD).

II PROCESS OVERVIEW

This overview will detail the steps of the A.C.E. reduction process. The steps listed below correspond to the steps in section VII, which contains programming instructions and is significantly less descriptive than this overview.

A. Read in the Sample Design File

The Sample Design File contains data for each block cluster selected in the first step of the listing sample. This file is the primary input file for the A.C.E. reduction since the sampling frame for the reduction is contained on the Sample Design File. Note that there are more clusters on the Sample Design File than were actually in the listing sample. This is due to the second step of listing sample selection, which was used to control expected listing workloads. The second step was needed only in Indiana and Missouri. The Current Sample Indicator (CSI) is used to screen out these clusters, which are not eligible for the A.C.E. reduction.

B. Assign Cluster Codes

Block clusters on the Sample Design File will be assigned two cluster codes for the A.C.E. reduction: Demographic strata codes and Consistency strata codes.

Demographic strata codes are based on the original Demographic/Tenure code assigned for listing sample selection (See Reference 4). The Demographic/Tenure code represents a classification of block clusters according to the approximate distribution of race/Hispanic origin and tenure, and was used as a sort variable in the selection of the listing sample. For the A.C.E. reduction, the 14 Demographic/Tenure codes will be combined into three Demographic strata, which are more than a sort variable in the reduction since sampling rates may vary across these strata. The three Demographic strata are:

- Minority: a block cluster with any minority (non-Other and non-Puerto Rico) Demographic/Tenure code
- Non-minority: a block cluster with any Other Demographic/Tenure code
- Puerto Rico: a block cluster with a Puerto Rico Demographic/Tenure code

Consistency strata codes are based on cluster HU count differences. An estimated HU count was created for listing sample selection based on the most recent Master Address File (MAF) HU counts available at the time. For the A.C.E. reduction, two updated cluster HU counts will be used: the preliminary independent listing (PIL) HU count and the Decennial Master Address File (DMAF) HU count. The PIL HU count is a preliminary HU count clerically tallied from the Independent Listing Book for each cluster in the listing sample. The DMAF HU count used in the A.C.E. reduction is taken from the January, 2000 version of the DMAF, which includes September and November Delivery Sequence File (DSF) updates. The PIL HU count and the DMAF HU count are compared and clusters are placed into consistency strata based on the relationship of those HU counts. Large differences between these counts indicate that

coverage problems may occur and thus the weights for such clusters should be controlled to avoid serious variance effects.

Clusters will be placed into three consistency strata:

- Low Inconsistent: a block cluster where the PIL HU count is more than 25 percent lower than the DMAF HU count. Low Inconsistent clusters may have a large percentage of erroneous enumerations in the census.
- Consistent: a block cluster where the absolute difference between the PIL HU count and the DMAF HU count is not more than 25 percent.
- High Inconsistent: a block cluster where the PIL HU count is more than 25 percent higher than the DMAF count. High Inconsistent clusters may have a large percentage of omissions in the census.

The percent differences to use as cutoffs to define the consistency strata are specified as the parameters X_L and X_H on the Reduction Parameter File.

For List/Enumerate clusters, the DMAF HU count will not be known at the time of the reduction. Thus, all such clusters will be classified as High Inconsistent.

C. Stratify Clusters

Clusters will be placed into A.C.E. reduction strata based on the Demographic and Consistency codes created in step B above and based on the collapsing pattern for a given state. Collapsing is required since some states do not have the sample to support a full crossing of Demographic and Consistency strata. The collapsing pattern will be predefined for each state on the Reduction Parameter File.

Medium stratum jumpers, those clusters that had been in the medium sampling stratum for the listing sample but now have 80 or more PIL HUs, have their own A.C.E. reduction stratum. Medium clusters were sampled at lower rates than large clusters in the listing sample since large clusters eventually undergo large block cluster subsampling, an operation that increases their weights. Medium stratum jumpers will also go through large block cluster subsampling, meaning the already high weights of these clusters will become even larger. By taking all or most of the medium stratum jumpers in the A.C.E. reduction, their weights will be controlled and these clusters will not introduce the significant weight variation they otherwise would have. Similarly, small block clusters that have 80 or more PIL HUs have their own A.C.E. reduction stratum to facilitate controlling their weights during large block cluster subsampling.

AIR and Puerto Rico block clusters will also be placed in their own A.C.E. reduction strata. Small block clusters that do not have 80 or more PIL HUs will

be assigned an A.C.E. reduction stratum code similar to medium and large clusters in order to compute reduction stratum target interview sample sizes for large block cluster subsampling.

For the complete set of reduction strata, see Attachment A.

D. Identify Eligible Clusters

Only certain block clusters are eligible for the A.C.E. reduction. The ineligible clusters will be removed from the process at this point. Information about these clusters, such as A.C.E. reduction stratum, will be saved to the Sample Design File for use in later A.C.E. sampling operations. The ineligible clusters are:

- Small block clusters
- AIR block clusters
- Puerto Rico block clusters

All other clusters will continue in the A.C.E. reduction process.

E. Calculation of Sampling Parameters

Prior to reduction, the Sample Design Team determined differential sampling rates for each reduction stratum relative to a baseline reduction stratum for each state. The differential sampling factors differ from reduction strata to reduction strata and from state to state depending on conditions such as the available sample in a state. The differential sampling factor for each stratum and state is provided on the Reduction Parameter File. The sample allocated to each A.C.E. reduction stratum accounts for the PIL HU count and the amount of differential sampling targeted for that stratum. If no differential sampling is desired in a state, the sample allocation is proportional to the PIL HU count in each stratum.

In addition to the different sampling rates among the A.C.E. reduction strata, medium and large clusters may be sampled at different rates. These different rates will occur in states where a listing adjustment or a second step of sampling was needed during the listing sample selection. The listing adjustment and the second step of sampling were two procedures built into the listing sample selection to control the expected number of HUs to list. These two procedures affected only large block clusters. Therefore, to restore the proportionality of the weights between the medium and large clusters, the take-everys for the large clusters will be adjusted to account for the listing adjustment and the second-step take-every.

Medium stratum jumpers have a predetermined take-every that is provided on the Reduction Parameter File. HUs in medium stratum jumper clusters are excluded

from the take-every calculations for the other A.C.E. reduction strata. It is likely that all medium stratum jumpers will be retained in the sample unless an overly large number of clusters are medium stratum jumpers. In this case, a subsample of medium stratum jumpers will be taken in the same manner used for subsampling all other eligible clusters.

In the reduction process, two take-everys are calculated: an initial take-every and a final take-every. The final take-every is calculated to facilitate variance estimation by providing an integer block cluster sample size for each reduction stratum and state.

F. Select a Subsample of Block Clusters

Following the calculation of the sampling parameters, a systematic sample of clusters is selected for the medium clusters and large clusters separately within each A.C.E. reduction stratum and state. Selected block clusters remain in sample and continue to the next sampling operation.

G. Update Files

After the sample selection, output files are updated or created. The first is the Sample Design File, which records results from each of the sampling operations for each block cluster. The second file is the housing unit sample size file, which contains initial target interview sample sizes for each reduction stratum.

III DEFINITIONS

A. States

All 50 states, the District of Columbia, and Puerto Rico are "states" in the A.C.E. Reduction.

B. Sampling Stratum

Block clusters are classified into four sampling strata for the listing sample selection based on an early census count of HUs used for clustering. These categories are as follows:

1. Small block clusters: 0 - 2 HUs.
2. Medium block clusters: 3 - 79 HUs and not on an AIR.
3. Large block clusters: 80 or more HUs and not on an AIR.
4. American Indian Reservation block clusters: 3 or more HUs and on an AIR.

The sampling strata above are the original sampling strata as located on the Sample Design File in the variable SS, Sampling Strata.

C. Preliminary Independent Listing HU Count

The independent listing HU count used in the A.C.E. reduction is preliminary because it is simply a clerical tally of HUs from the independent listing books. These A.C.E. HU counts are obtained from the Cluster Count File, a file provided to the Decennial Systems and Contracts Management Office (DSCMO) from the Technologies Management Office (TMO). See Reference 8 for the specifications on independent listing file transfers.

D. Decennial Master Address File HU Counts

Census HU counts are obtained from the DMAF. The variables GQFLG (Group Quarters HU Flag) and SMAFID (Surviving MAFID) from the DMAF represent the characteristics of an address used in identifying DMAF HUs for the purposes of the A.C.E. reduction. The GQFLG code distinguishes Census HUs from group quarters. The SMAFID code identifies duplicate HUs on the DMAF. Possible values for GQFLG and SMAFID are as follows:

GQFLG: 0 = Housing Unit
1 = Special Place
2 = Group Quarters
3 = GQ Embedded Housing Unit

SMAFID: 0 = address is not a duplicate. (A non-zero SMAFID implies the address is a duplicate.)

The GQFLG and SMAFID codes that will be recognized as valid DMAF HUs for A.C.E. reduction are as follows:

GQFLG = 0 and SMAFID = 0, or
GQFLG = 3 and SMAFID = 0.

E. Consistency Strata

Consistency strata are groups of clusters formed on the basis of the percent difference between the PIL HU count and the DMAF HU count. A cluster is allocated to a consistency stratum by the definitions in Table 1:

Table 1. Consistency Strata Definitions

Consistency Stratum	If	Then
	Criteria	Consistency Code
Low Inconsistent	$PIL < X_L \times DMAF$	1
Consistent	$X_L \times DMAF \leq PIL \leq X_H \times DMAF$	2
High Inconsistent	$PIL > X_H \times DMAF$	3

The variables X_H and X_L are "Inconsistency Cutoffs." X_H is the high inconsistency cutoff ($X_H = 1.25$ meaning 25 percent higher than the PIL) and X_L is the low inconsistency cutoff ($X_L = 0.75$ meaning 25 percent lower than the PIL). These cutoffs are specified on the Reduction Parameter File. At the time of A.C.E. reduction, List/Enumerate clusters do not have a DMAF HU count. Hence, all List/Enumerate clusters are in the High Inconsistent Stratum.

F. Demographic Strata

Demographic strata are groups of clusters formed on the basis of the estimated 1990 racial and Hispanic ethnicity distribution. Using the Demographic/Tenure Group Code assigned during the listing sample selection, labeled DTCODE on the Sample Design File, a cluster is allocated to a demographic stratum based on the definition in Table 2:

Table 2. Demographic Strata Definitions

Demographic Stratum	IF	THEN
	Criteria	Demog. Strata Code
Minority	DTCODE = 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10	1
Non-minority	DTCODE = 11 or 12	2
Puerto Rico	DTCODE = 13 or 14	3

G. A.C.E. Reduction Strata

Block clusters are classified into one of nineteen (19) A.C.E. reduction strata based on original sampling stratum code, demographic stratum code, consistency stratum code, and a collapsing flag. See Attachment A for a complete list of all nineteen A.C.E. reduction strata.

H. Stratum Jumpers

Between the time of selecting the listing sample and the A.C.E. reduction, it is possible for clusters to change from their original sampling stratum. Such clusters are referred to as Stratum Jumpers. The only stratum jumpers of interest at this point, however, are the medium stratum to large stratum jumpers because these clusters require special attention in the reduction to control their weights. Medium stratum jumpers are clusters that are in the medium sampling stratum on the Sample Design File but have a PIL HU count of 80 or more. Small stratum jumpers, those clusters in the small sampling stratum at the time of listing but with a PIL HU count of 80 or more, are also identified in A.C.E reduction but are dealt with during small block cluster subsampling and large block cluster subsampling.

I. A.C.E. Reduction Parameters

The DSSD will provide certain parameters needed for the A.C.E. reduction on the Reduction Parameter File (See Attachment B for a layout).

1. Differential Sampling Factors

The differential sampling factors are the sampling rates relative to the baseline A.C.E. reduction stratum. The baseline A.C.E. reduction stratum is the stratum with the lowest sampling rate. In every state, the Consistent stratum is the baseline. The differential sampling factors indicate the degree to which the Minority, Inconsistent Low, and Inconsistent High reduction strata are differentially sampled relative to the Non-Minority Consistent reduction stratum. For example, a differential sampling factor of two for the Minority reduction stratum means that the probability of selection for a cluster in the Minority stratum is twice that of a cluster in the Consistent stratum.

2. Inconsistency Cutoffs

The Inconsistency Cutoffs are the critical values (in terms of percentages) that define significant differences between the PIL and DMAF HU counts. For the purposes of the reduction, a significant difference is more than 25 percent. Significant differences can occur on both the high end and the low end, hence the parameters X_H and X_L.

3. Listing Adjustment

The listing adjustment is a factor that was applied to increase the first-step take-every of the large sampling stratum during listing sample selection to reduce the expected listing workload in some states. The A.C.E. reduction will compensate for the decrease of large clusters in the listing sample due to the listing adjustment.

4. Second-step Take-every

During listing sample selection, a second step of sampling was required in some states in the large sampling stratum to reduce the expected listing workload. The A.C.E. reduction will compensate for the second-step sampling of large clusters.

5. Collapsing Flag

The collapsing flag is a flag on the Reduction Parameter File that indicates the prespecified collapsing pattern that will be used to form the A.C.E. reduction strata in each state. The collapsing flag represents a strategy to assign an A.C.E. reduction stratum to a block cluster based on its original sampling stratum and demographic/consistency characteristics. An example of the use of the collapsing flag is illustrated on page 17.

6. Medium and Large Block Cluster Weights after Listing Sample Selection

Large clusters were sampled at a higher rate than medium clusters during the listing sample selection, creating different weights for medium and large clusters. These weights will be required to calculate reduction sampling parameters.

IV ASSUMPTIONS

A. The A.C.E. listing sample of block clusters was selected according to the previously planned ICM 750,000 HU design. There are 29,695 block clusters in the listing sample including 559 in Puerto Rico.

B. Independent listing HU counts are preliminary. This is due to time constraints. Any reference to the Independent List at this stage in the A.C.E. sample survey will be referred to as the Preliminary Independent List (PIL). Post-A.C.E. reduction processes such as small block cluster subsampling and large block cluster subsampling will have a "Keyed and Valid" independent listing of HUs. "Keyed and Valid" implies these HU counts will have undergone a complete quality control.

C. The final A.C.E. sample size is approximately 300,000 HUs. This sample size is for interview after the A.C.E. reduction, small block cluster subsampling, and large block cluster subsampling.

D. Only medium and large block clusters are sampled in the A.C.E. Reduction. All small block clusters, American Indian Reservation block clusters, and Puerto Rico block clusters remain in the sample.

E. A.C.E. reduction stratum codes are stored in two-digit fields and are zero-filled as needed.

F. Decimal numbers are rounded to six (6) digits unless otherwise noted at the time of creation using the standard rounding procedure. Standard rounding in this specification means that a number with a seventh-decimal value of five or higher is rounded up in the sixth decimal. Otherwise, the sixth decimal value is unchanged.

V INPUT FILES

The following files will be used in this process.

A. Reduction Parameter File

This file contains the parameters for each state that are needed for the A.C.E. reduction. This file will be provided by the DSSD. There is one record for each A.C.E. reduction stratum within each state. See Attachment B for the file layout and an example of the data in the Reduction Parameter File.

B. Decennial Master Address File (DMAF)

This file contains address information for each HU in Census 2000. The DMAF is formed from an extract of the MAF along with updates from the United States Postal Service DSF and from census operations such as Local Update of Census Addresses. See Reference 9 for more information on the DMAF.

C. Cluster Count File

The Cluster Count File contains one record for each of the 29,695 block clusters in the A.C.E. listing sample. The TMO will transmit a file to the DSCMO at the end of the independent listing operation. Each record will contain a preliminary count of the HUs listed in each cluster during independent listing.

Attachment C contains the layout of this file. For further details on this file, see Reference 8.

D. Sample Design File

The Sample Design File contains one record per block cluster chosen during the first step of listing sample selection. This file tracks the path that each block cluster travels during the A.C.E. sampling procedures. The Sample Design File contains categorical variables corresponding to each procedure as well as parameters and HU totals. If the block cluster fell out of sample at some point, the remaining variables are left blank. The variable CSI is used to indicate which

block clusters are in sample; clusters with a CSI of one are in the sample. The initial version of the file, which was created following the listing sample selection and is the input for the A.C.E. reduction, is called SDF.US1. There are 29,717 records on the Sample Design File. Attachment D contains a file layout for the Sample Design File.

VI OUTPUT FILES

A. Housing Unit Sample Size File

The HU Sample Size File contains three variables for large block cluster subsampling: State, A.C.E. Reduction Stratum, and Target Number of HUs.

Variable Description	Name	Location
State code	STATE	1-2
A.C.E. Reduction Stratum (zero-filled)	ARST	4-5
Target Number of HUs to interview in A.C.E. reduction stratum	T	7-14

B. Sample Design File

Updates will be made to the Sample Design File based on the results of the A.C.E. reduction. After all states have been verified, the new version of the Sample Design File will be called SDF.US2.

C. Reduction Parameter File

Updates will be made to the Reduction Parameter File during the A.C.E. reduction process so that the Sample Design Team in the DSSD may check the parameters for statistical validity.

VII REDUCTION PROCESS

Process each state as follows:

A. Read in Sample Design File

The layout for the Sample Design File is located in Attachment D. In the current state, read in the following fields for each cluster with CSI = 1:

Variable Description	Name	Location
State code	STATE	3-4
Current Sample Indicator	CSI	19
A.C.E. block cluster number and check digit	CLUST	21-26
List/Enumerate Indicator	LEIND	33
Sampling Stratum	SS	55
Demographic/Tenure group code	DTCODE	57-58
First-step index number	INDEX1	92-99
Unbiased weight after listing sample	WEIGHTBC	153-164

B. Assign Cluster Codes

1. Demographic Strata Codes

Assign the demographic stratum code created below to each cluster with CSI = 1 on the Sample Design File. Using DTCODE from the Sample Design File, assign DEMCODE to each cluster using the rules in Table 3:

Table 3. Demographic Stratum Code Assignment Rules

Demographic Stratum	IF	THEN
	Criteria	DEMCODE
Minority	DTCODE = 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10	1
Non-minority	DTCODE = 11 or 12	2
Puerto Rico	DTCODE = 13 or 14	3

2. Consistency Strata Codes

Assign a consistency stratum code to each cluster with CSI = 1 using the following steps:

a. Obtain the following information for each cluster:

- the DMAF HU count from the January, 2000 DMAF as defined in section V.B, NHUDMAF
- the DMAF HU count from the December, 1999 DMAF, NHUDMAFI (See Reference 7 for more information on the two DMAF versions.)
- the PIL HU count from the Cluster Count File, NHUILP
- the Inconsistency Cutoffs for the current state from the Reduction Parameter File, X_H and X_L.

b. For each List/Enumerate (L/E) cluster on the Sample Design File, assign a consistency stratum code, CONCODE, equal to "3", High Inconsistent. The L/E clusters have LEIND = 1 on the Sample Design File.

c. For each non-L/E cluster, assign a Consistency Stratum Code, CONCODE, using the PIL HU count, the DMAF HU count, and the Inconsistency Cutoffs according to the rules in Table 4:

Table 4. Consistency Stratum Code Assignment Rules

Consistency Stratum	If	Then
	Criteria	CONCODE
Low Inconsistent	$PIL < X_L \times DMAF$	1
Consistent	$X_L \times DMAF \leq PIL \leq X_H \times DMAF$	2
High Inconsistent	$PIL > X_H \times DMAF$	3

C. Stratify Clusters

Assign each cluster with CSI = 1 on the Sample Design File to an A.C.E. reduction stratum using the following hierarchical rules:

1. Assign all AIR clusters an A.C.E. reduction stratum of "17". These are clusters in the AIR sampling stratum (SS = 4) on the Sample Design File.

2. Assign all medium stratum jumpers an A.C.E. reduction stratum of "16". Medium stratum jumpers are clusters that were originally in the medium sampling stratum (SS = 2) for the listing sample but have a PIL HU count of 80 or more.

3. Assign all small stratum jumpers an A.C.E. reduction stratum of "19". Small stratum jumpers are clusters that were originally in the small sampling stratum (SS = 1) for the listing sample but have a PIL HU count of 80 or more.

4. Assign all Puerto Rico clusters that are not medium or small stratum jumpers an A.C.E. reduction stratum of "18". These are clusters having a FIPS state code (STATE) of "72".

5. Obtain the Collapsing Flag for the current state from the Reduction Parameter File.

6. Assign A.C.E. reduction strata codes to all other clusters using Table 5 below:

Table 5. A.C.E. Reduction Strata Assignments

DEMCODE	CONCODE	Collapsing Flag								
		01	02	03	04	05	06	07	08	09
1	1	01	01	01	07	10	11	11	13	14
1	2	01	01	01	08	08	12	12	13	12
1	3	01	01	01	09	10	11	11	13	15
2	1	02	05	06	07	10	02	05	13	02
2	2	03	03	06	08	08	03	03	13	03
2	3	04	05	06	09	10	04	05	13	04

The strategy of Table 5 is, given a collapsing flag for a state, work down the column of the collapsing flag until the appropriate DEMCODE and CONCODE are reached and assign the cluster to the A.C.E. reduction stratum in the corresponding cell of the table.

As an illustration, given a cluster with a DEMCODE of "2" implying a non-minority cluster, a CONCODE of "3" meaning a High Inconsistent cluster, and a collapsing flag of "05" for the state, then the A.C.E. Reduction Stratum is "10." Using Attachment A, an A.C.E. reduction stratum of "10" is called "Inconsistent". Therefore, this A.C.E. reduction stratum was collapsed over demographics (minority and non-minority) and over inconsistency (high and low) to include all "Inconsistent" clusters.

D. Identify Eligible Clusters

At this point, clusters will be identified as eligible or ineligible for further processing in the A.C.E. reduction. Small block clusters, AIR block clusters, and Puerto Rico block clusters are ineligible, while all other clusters are eligible.

1. For all block clusters with SS = 1, SS = 4, or STATE = 72, update the following variables on the Sample Design File. Set TEAR = 1.000000, RSAR = 1.000000, ACERED = 1, WEIGHTAR = WEIGHTBC,

INDEXR = " ", and COLFLAG = the collapsing flag for that state from the Reduction Parameter File.

Variable Description	Name	Location
Preliminary Number of HUs on the Independent List	NHUILP	176-180
Number of HUs on the January 2000 DMAF	NHUDMAF	182-186
Demographic Code	DEMCODE	188
Consistency Code	CONCODE	189
A.C.E. Reduction Stratum (zero-filled)	ARST	190-191
A.C.E. Reduction Indicator	ACERED	193
Random Start for A.C.E. Reduction	RSAR	195-205
Take-every for A.C.E. Reduction	TEAR	207-217
Unbiased weight after A.C.E. reduction	WEIGHTAR	219-230
Collapsing Flag	COLFLAG	232
A.C.E. Reduction Index Number	INDEXR	234-241
Number of HUs on the December 1999 DMAF (Initial)	NHUDMAFI	243-247

For ineligible clusters, the A.C.E. reduction process terminates at this point.

2. For all other block clusters, continue to step E below.

E. Calculation of Sampling Parameters

1. Obtain the Variables for Calculating Take-Everys and Expected Sample Sizes

   a. Obtain the following information for the current state from the A.C.E. Reduction Parameter File:

      i. Differential sampling factors for each A.C.E. reduction stratum, K
      ii. Listing Adjustment, LISTADJ
      iii. Second step take-every, TE2
      iv. Medium stratum jumper take-every, TESJ
      v. Medium cluster weight after the listing sample selection, WTM
      vi. Large cluster weight after the listing sample selection, WTL
      vii. Target number of housing units for interview in the current state, THU. Note: THU does not include HUs in stratum jumper clusters.

   b. Obtain the Sampling Stratum, SS, for each eligible cluster from the Sample Design File.

c. Obtain the PIL HU count, NHUILP for each eligible cluster from the Cluster Count File.

d. Calculate the following tallies for both medium (SS = 2) and large (SS = 3) clusters in each A.C.E. reduction stratum in the current state:

i. Total number of medium clusters in the listing sample, in the i^{th} A.C.E. reduction stratum, $NCLUSTM_i$
ii. Total number of large clusters in the listing sample, in the i^{th} A.C.E. reduction stratum, $NCLUSTL_i$
iii. Preliminary Listing HU count within medium clusters, in the i^{th} A.C.E. reduction stratum, $PILM_i$
iv. Preliminary Listing HU count within large clusters, in the i^{th} A.C.E. reduction stratum, $PILL_i$

2. Calculate the Initial Take-Everys

a. Calculate the target number of interview HUs for the i^{th} A.C.E. reduction stratum.

For reduction strata 1 - 15:

$$T_i = THU \times \frac{K_i \times \left(PILM_i + \frac{WTL}{WTM} PILL_i \right)}{\sum_{i \in (1 to 15)} K_i \times \left(PILM_i + \frac{WTL}{WTM} PILL_i \right)}$$

For reduction stratum 16:

$T_{16} = 1.0$.

b. Calculate the initial take-every for the medium clusters in the i^{th} A.C.E. reduction stratum.

For reduction strata 1 - 15:

$$ITEM_i = \frac{PILM_i + \frac{WTL}{WTM} \times PILL_i}{T_i}$$

If $ITEM_i < 1$, then output $ITEM_i$ to the updated Reduction Parameter File but continue the A.C.E. reduction process with $ITEM_i = 1.0$.

For reduction stratum 16:

$ITEM_{16}$ = TESJ as provided on the Reduction Parameter File.

c. Calculate the initial take-every for the large clusters in the i^{th} A.C.E. reduction stratum.

For reduction strata 1 - 15:

$$ITEL_i = ITEM_i \times \frac{LISTADJ}{TE2}$$

If $ITEL_i < 1.0$, then set $ITEL_i = 1.0$.

For convenience of future computations, set $ITEL_{16} = 1.0$ since stratum jumpers are either in the medium sampling stratum or the small sampling stratum and not in the large stratum.

3. Calculate the expected HU interview sample size and expected cluster interview sample size for both medium and large clusters in the i^{th} reduction stratum, including reduction stratum 16.

$$EHUM_i = \frac{PILM_i}{ITEM_i} \qquad ECLUSTM_i = \frac{NCLUSTM_i}{ITEM_i}$$

$$EHUL_i = \frac{PILL_i}{ITEL_i} \qquad ECLUSTL_i = \frac{NCLUSTL_i}{ITEL_i}$$

For stratum 16, let $EHUL_{16} = 0$ and $ECLUSTL_{16} = 0$.

Round $ECLUSTM_i$ and $ECLUSTL_i$ to integers using standard rounding. Denote these rounded numbers as follows:

$$RECLUSTM_i = int(ECLUSTM_i + 0.5)$$
$$RECLUSTL_i = int(ECLUSTL_i + 0.5)$$

If $NCLUSTM_i > 0$, then

If $RECLUSTM_i = 0$, then set $RECLUSTM_i = 1$ to ensure that one cluster is sampled in the stratum and continue to Step 4a;
If $RECLUSTM_i \neq 0$, then continue to Step 4a.

If $NCLUSTL_i > 0$, then

If $RECLUSTL_i = 0$, then set $RECLUSTL_i = 1$ to ensure that one cluster is sampled in the stratum and continue to Step 4c;
If $RECLUSTL_i \neq 0$, then continue to Step 4c.

4. Calculate the Final Take-Everys and Random Starts

a. Calculate the final take-every for the medium clusters in the i^{th} A.C.E. reduction stratum, including reduction stratum 16.

$$TEM_i = \frac{NCLUSTM_i}{RECLUSTM_i}$$

If $NCLUSTM_i = 0$, then set $TEM_i = 0$.

b. Generate a random number, RNM, between 0 and 1 ($0 < RNM \leq 1$), and calculate the random start for medium clusters in the i^{th} A.C.E. reduction stratum, including stratum 16 (medium stratum jumpers). Generate a new random number for each A.C.E. reduction stratum and each state.

$$RSM_i = TEM_i \times RNM$$

c. Calculate the final take-every for the large clusters in the i^{th} A.C.E. reduction stratum.

$$TEL_i = \frac{NCLUSTL_i}{RECLUSTL_i}$$

Set $TEL_{16} = 1.0$ since stratum jumpers are either in the medium sampling stratum or the small sampling stratum. TEL_{16} does not apply but is set for completeness.

If $NCLUSTL_i = 0$, then set $TEL_i = 0$.

d. Generate a random number, RNL, between 0 and 1 ($0 < RNL \leq 1$), and calculate the random start for large clusters in the i^{th} A.C.E. reduction stratum. Generate a new random number for each A.C.E. reduction stratum and each state.

$$RSL_i = TEL_i \times RNL$$

5. Calculate the cluster weight for medium and large clusters after the A.C.E. block cluster reduction for the i^{th} reduction stratum, including reduction stratum 16.

$$WTARM_i = WTM \times TEM_i$$
$$WTARL_i = WTL \times TEL_i$$

6. Update the Reduction Parameter File. The parameter file has one record for each reduction stratum per state. For each of the reduction strata, update the following variables:

Variable Description	Name	Location
Total medium clusters after the listing sample selection	NCLUSTM	86-90
Total large clusters after the listing sample selection	NCLUSTL	92-96
Preliminary Indep. Listing HU Count in medium clusters	PILM	98-103
Preliminary Indep. Listing HU Count in large clusters	PILL	105-110
Target interview sample size for the A.C.E. reduction stratum	T	111-120
Take-every for medium clusters	TEM	121-130
Take-every for large clusters	TEL	132-141
Random start for medium clusters	RSM	143-152
Random start for large clusters	RSL	154-163
Random number for medium clusters	RNM	165-172
Random number for large clusters	RNL	174-181
Expected number of housing units in medium clusters	EHUM	183-188
Expected number of housing units in large clusters	EHUL	190-195
Expected number of medium clusters	ECLUSTM	197-202
Expected number of large clusters	ECLUSTL	204-209
Medium cluster weight following A.C.E. reduction	WTARM	211-221
Large cluster weight following A.C.E. reduction	WTARL	223-233
Initial take-every for medium clusters	ITEM	234-242
Initial take-every for large clusters	ITEL	243-251

7. Provide the Reduction Parameter File to the Sample Design Team in the DSSD for review. If the calculation of the take-everys results in some values less than 1, then the differential sampling factors may need to be revised and the parameters recalculated. Wait for approval of the sampling parameters before proceeding to section F.

F. Select a Subsample of Block Clusters

For each of the A.C.E. reduction strata crossed with the original sampling strata, medium and large, select a separate systematic sample of block clusters as follows:

1. Sort the block clusters in the following order:

   - Sampling Stratum (SS).
   - A.C.E. Reduction Stratum (ARST).
   - Consistency Stratum (CONCODE).
   - List/Enumerate Indicator (LEIND).
   - Index Number (INDEX1) on the Sample Design File.

2. Assign an order number to each cluster in the sampling stratum and A.C.E. reduction stratum currently being subsampled. Give the first cluster in the sort an order number of "1", and increment by one for all remaining clusters. The assigned number is referred to as the A.C.E Reduction Index Number. Place the A.C.E. Reduction Index Number (INDEXR) on the Sample Design File.

3. Generate a sequence of numbers $L_1, ..., L_n$ as follows:

   - Obtain the Random Start (RSAR) and the Take-every (TEAR) for A.C.E. Reduction. If the current sampling stratum is medium (SS = 2), set RSAR = RSM and TEAR = TEM, where RSM and TEM are obtained from the Reduction Parameter File. If the current sampling stratum is large (SS = 3), set RSAR = RSL and TEAR = TEL, where RSL and TEL are obtained from the Reduction Parameter File.

   - Let L_1 = RSAR

   - Calculate $L_j = L_{j-1} + TEAR$, for j = 2 to n where n is the largest integer such that $[RSAR + (n - 1) \times TEAR] \leq N$, where N is the largest order number in the sampling stratum and A.C.E. reduction stratum currently being subsampled.

   - Round each L_j up to the nearest integer (an integer rounds to itself).

- For each block cluster in the sampling stratum and the A.C.E. reduction stratum:

  If the order number is equal to the rounded values of L_j, j = 1, ..., n, then do the following:

  - Assign the A.C.E. Reduction Indicator (ACERED) on the Sample Design File equal to "1". The block cluster was selected in sample.

  - Calculate the block cluster weight (WEIGHTAR) following the A.C.E. Reduction. Obtain the Take-every for listing sample selection, TE1, and the second-step Take-every, TE2, from the Sample Design File, and compute the weight as follows;

    $$WEIGHTAR = TE1 \times TE2 \times TEAR$$

  If the order number does not equal any of the rounded values of L_j, j = 1, ..., n, then do the following:

  - Assign the A.C.E. Reduction Indicator (ACERED) on the Sample Design File equal to "0". The block cluster was not selected in sample.

  - Set the Current Sample Indicator (CSI) on the Sample Design File equal to "0". The block cluster was not selected so it is not currently in sample.

- For example: if N = 100, RSAR = 2.4 and TEAR = 7.2, then n = 14. Set $L_1 = 2.4$. The generated L_js would be the sequence: 2.4, 9.6, 16.8, 24.0, ..., 96.0. Therefore, the block clusters with ordered numbers 3, 10, 17, 24, 32, ..., and 96 would be selected for the sample.

4. Compute a Check

For each reduction stratum, check the number of sampled block clusters, given by n, by calculating c:

$$c = \left| \frac{N}{TEAR} - n \right|$$

If the sampling is implemented correctly, c will be less than 1. For values of c that are not less than one and have not been resolved, contact the DSSD for review of the sampling operations.

G. Update and Create Files

1. Update the Sample Design File. This file tracks the path that each sampled block cluster travels during the A.C.E. sampling procedures. It was created following the listing sample selection and contains one record per block cluster selected during the listing sample selection. Version 2 will be created by updating version 1 with the A.C.E. reduction information. The file layout is in Attachment D. Update the file with the following A.C.E. block cluster reduction information:

Variable Description	Name	Location
Preliminary Number of HUs on the Independent List	NHUILP	176-180
Number of HUs on the January 2000 DMAF	NHUDMAF	182-186
Demographic Code	DEMCODE	188
Consistency Code	CONCODE	189
A.C.E. Reduction Stratum (zero-filled)	ARST	190-191
A.C.E. Reduction Indicator	ACERED	193
Random Start for A.C.E. Reduction	RSAR	195-205
Take-every for A.C.E. Reduction	TEAR	207-217
Unbiased weight after A.C.E. reduction	WEIGHTAR	219-230
Collapsing Flag	COLFLAG	232
A.C.E. Reduction Index Number	INDEXR	234-241
Number of HUs on the December 1999 DMAF (Initial)	NHUDMAFI	243-247

2. Housing Unit Sample Size File. This file contains three variables which are the basis of the input to large block cluster subsampling. Create one record for each reduction stratum per state which includes the following variables:

Variable Description	Name	Location
State code	STATE	1-2
A.C.E. Reduction Stratum (zero-filled)	ARST	4-5
Target Number of Housing Units to Interview in A.C.E. Reduction Stratum	T	7-14

VIII REFERENCES

1 DSSD Census 2000 Procedures and Operations Memorandum Series R-8, "Census 2000 Specifications for Block Cluster Formation-Reissue," May 3, 1999.

2 DSSD Census 2000 Procedures and Operations Memorandum Series R-9, "Amendment to Census 2000 Specifications for Block Cluster Formation–Reissue," May 3, 1999.

3 DSSD Census 2000 Procedures and Operations Memorandum Series R-10, "Accuracy and Coverage Evaluation (ACE) Survey: Second Amendment to Census 2000 Specifications for Block Cluster Formation–Reissue," May 3, 1999.

4 DSSD Census 2000 Procedures and Operations Memorandum Series R-5, "Accuracy and Coverage Evaluation Survey: Universe File and Block Cluster Sampling Parameter File Specification," March 30, 1999.

5 DSSD Census 2000 Procedures and Operations Memorandum Series R-3, "Accuracy and Coverage Evaluation Survey: Block Cluster Sample Selection Specification," March 29, 1999.

6 DSSD Census 2000 Procedures and Operations Memorandum Series R-23, "Accuracy and Coverage Evaluation Survey: Approval and Summary of Results of the Reduction Sample," January 21, 2000.

7 DSSD Census 2000 Procedures and Operations Memorandum Series R-22, "Accuracy and Coverage Evaluation Survey: Cluster Reduction Contingency Plan," December 16, 1999.

8 DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-FA-02 - Revision #1, TMO A.C.E. 2000 Planning Memorandum Series #2, "Revision #1 of A.C.E. 2000 Independent Listing File Transfers (Draft)," July 29, 1999–DRAFT

9 DSSD Census 2000 Procedures and Operations Memorandum Series D-1, "Specification of the Decennial Master Address File Deliverability Criteria for Census 2000," June 30, 1999.

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
A.C.E. Implementation Team
Statistical Design Team Leaders
DSSD Sample Design Team

A.C.E. Reduction Strata

Code	Stratum Name
01	Minority
02	Non-minority Low Inconsistent
03	Non-minority Consistent
04	Non-minority High Inconsistent
05	Non-minority Inconsistent
06	Non-minority
07	Low Inconsistent
08	Consistent
09	High Inconsistent
10	Inconsistent
11	Minority Inconsistent
12	Minority Consistent
13	Full Collapse
14	Minority Low Inconsistent
15	Minority High Inconsistent
16	Medium Stratum Jumpers
17	American Indian Reservations
18	Puerto Rico
19	Small Stratum Jumpers

A.C.E. Reduction Parameter File Layout and Example

Variable Description	Name	Location	Format
FIPS State Code	STATE	1-2	I2
A.C.E. Reduction Stratum	ARST	4-5	I2
State Target Number of HUs to Interview	THU	7-11	I5
Listing Adjustment for State	LISTADJ	13-20	F8.4
Second-step Take-every for state	TE2	22-29	F8.4
Medium Stratum Jumper Take-every	TESJ	31-38	F8.4
Low Inconsistency Cutoff	XL	40-43	F4.2
High Inconsistency Cutoff	XH	45-48	F4.2
Collapsing Flag	COLFLAG	50-51	I2
Differential Sampling Factor	K	53-61	F9.4
Medium cluster weight after the listing sample selection	WTM	63-72	F10.4
Large cluster weight after the listing sample selection	WTL	74-83	F10.4

Example of Reduction Parameter File for Alabama (State Code = 1):

1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0
01 01 4470	1.0000	1.0000	1.0000	0.75	1.25 1	2.0000	168.8000	34.4000
01 02 4470	1.0000	1.0000	1.0000	0.75	1.25 1	1.5000	168.8000	34.4000
01 03 4470	1.0000	1.0000	1.0000	0.75	1.25 1	1.0000	168.8000	34.4000
01 04 4470	1.0000	1.0000	1.0000	0.75	1.25 1	1.7500	168.8000	34.4000

In this example, the state total number of HUs to interview is 4,470, and both the listing adjustment and second-step take-every are 1. The Take-every for medium stratum jumpers is also 1. The cutoff for defining Low Inconsistency clusters is 0.75, meaning the PIL HU count must be at least 25 percent lower than the DMAF HU count for a cluster to be in the Low Inconsistent stratum. Likewise, the High Inconsistent Cutoff is 1.25, so the PIL must be at least 25 percent higher than the DMAF to be a High Inconsistent cluster. The collapsing flag is 1, meaning all minority clusters are in the same A.C.E. reduction stratum, while the non-minority clusters remain split into the three consistency strata, resulting in four A.C.E. reduction strata in Alabama and thus the four records in the example (See Table 5 on page 17). The A.C.E. reduction stratum codes are given in the second field. The differential sampling factor for each stratum is in the tenth field. The differential sampling factors give an indication of the differential sampling that occurs in Alabama in this example. Minority clusters will be sampled at twice the rate of non-minority consistent clusters, so the take-every for minority clusters is half the take-every of non-minority consistent clusters. Similarly, the take-every for non-minority low inconsistent clusters is 2/3 that of the non-minority consistent clusters. The medium cluster weight following the listing sample selection for Alabama is 168.8 and the large cluster weight is 34.4.

It is important to note that this is an example for illustrative purposes. It is very likely that the production parameters for Alabama will differ from those in this example.

Cluster Count File Layout

This file will be provided by the TMO.

Variable Description	Name	Location
LCO Number (digits 1-2 are RO code)	LCO	1-4
FIPS State Code	ST	5-6
FIPS County Code	CC	7-9
Cluster Number (with check digit)	CLUSTER	10-15
Preliminary number of Independent Listing HUs	NHUILP	16-20

Sample Design File Layout

Variable Description	Name	Places	Source
Census Region	REGION	1	UN
Census Division	DIV	2	UN
State code	STATE	3-4	UN
County code	COUNTY	5-7	UN
Local census office	LCO	8-11	CS
Interim Tract (Pseudo Tract)	ITRACT	12-17	BC
Current Sample Indicator	CSI	19	UO
A.C.E. block cluster number	CLUST	21-25	CS
Check Digit	DIGIT	26	CS
Geography block cluster number	GCLUST	28-32	BC
List/Enumerate Indicator	LEIND	33	BC
Type of Enumeration Area Recode	TEACR	34	CS
Type of Enumeration Area group	TEAG	36	BC
Number of HUs used for sample design	NHU	37-41	BC
Number of MAF HUs	NHUM	43-47	BC
Number of 1990 HUs	NHU90	49-53	BC
Sampling Stratum 1 = Small 2 = Medium 3 = Large 4 = American Indian Reservation	SS	55	UN
American Indian Country Indicator 0 = No American Indian Country 1 = American Indian Reservation/trust land 2 = Tribal Jurisdiction Area/ Alaska Native Village Statistical Area/ Tribal Designated Statistical Area	AICIND	56	BC
Demographic/Tenure Group code	DTCODE	57-58	UN
Demographic/Tenure Group label	DTLABEL	59-60	UN
Estimated Urbanicity of block cluster 1 = Urban Area with population ≥250,000 2 = Other Urban Area 3 = Non-Urban Area	ECLUSURB	62	UN
Size Category 1 = Small (0-2 hus) 2 = Medium (3-79 hus) 3 = Large (80+ hus)	SIZCAT	63	UN
Additional space		64-91	
First step index number	INDEX1	92-99	CS
Listing sample selection Indicator 1 = Selected	BC1	101	CS
Random Start for listing sample selection	RS1	103-113	UN
Take-every for listing sample selection	TE1	115-125	UN
Second block cluster sampling Indicator 0 = Not Selected 1 = Selected	BC2	127	CS

Random Start for second block cluster sampling	RS2	129-139	CS
Take-every for second block cluster sampling	TE2	141-151	CS
Unbiased weight after block cluster sampling	WEIGHTBC	153-164	CS
Additional space		165-175	
Preliminary Number of HUs on the Independent List	NHUILP	176-180	AR
Number of Housing Units on the January 2000 DMAF	NHUDMAF	182-186	AR
Demographic Code 1 = Minority 2 = Non-minority 3 = Puerto Rico	DEMCODE	188	AR
Consistency Code 1 = Low Inconsistent (PIL significantly smaller than DMAF) 2 = Consistent 3 = High Inconsistent (PIL significantly larger than DMAF)	CONCODE	189	AR
A.C.E. Reduction Stratum (zero-filled)	ARST	190-191	AR
A.C.E. Reduction Indicator 0 = Not Selected 1 = Selected	ACERED	193	AR
Random Start for A.C.E. Reduction	RSAR	195-205	AR
Take-every for A.C.E. Reduction	TEAR	207-217	AR
Unbiased weight after A.C.E. reduction	WEIGHTAR	219-230	AR
Collapsing Flag	COLFLAG	232	AR
A.C.E. Reduction Index Number	INDEXR	234-241	AR
Number of Housing Units on the December 1999 DMAF (Initial)	NHUDMAFI	243-247	AR
Additional space		248-300	

Source Codes

AR: A.C.E. Reduction
BC: Block Clustering
CS: Block Cluster Sampling
UN: Universe File Creation
UO: Updated for each operation